

科学研究費助成事業 研究成果報告書

平成 30 年 6 月 2 日現在

機関番号：13302

研究種目：挑戦的萌芽研究

研究期間：2016～2017

課題番号：16K12458

研究課題名(和文) 音声伝送指標を利用した音声プライバシー保護

研究課題名(英文) Study on method for protecting speech privacy using speech transmission index

研究代表者

鷓木 祐史 (UNOKI, MASASHI)

北陸先端科学技術大学院大学・先端科学技術研究科・教授

研究者番号：00343187

交付決定額(研究期間全体)：(直接経費) 2,600,000円

研究成果の概要(和文)：音声プライバシー保護は、オープンスペースにおける音声情報漏えいを防ぐ意味で挑戦的な課題である。しかし、会話音声を不明瞭で聴き取り難くすることはとても難しい。本研究では、漏えい音声を聴取してしまう人の音環境を、後処理した漏えい音声を呈示することで所望の音声伝送指標を得られるように制御し、漏えい音声の明瞭性・聴き取りにくさを能動的に制御する。聴取実験による評価の結果、提案法はSTIを能動的に制御することで漏えい音声の明瞭性を低下させ音声プライバシーを保護可能であることがわかった。また、従来法(マスクング法と残響付与法)と比較して、効果的にプライバシー保護を実現できることもわかった。

研究成果の概要(英文)：Protecting speech privacy is an important challenge in room acoustics. However, protecting people's conversation from being overheard by an unintended listener, that is, making them not understandable, is difficult. This research study a method for protecting speech privacy by actively controlling the speech transmission index (STI). In this method, the STI in the simulated room can be controlled by manipulating the parameters of the simulated room impulse response (RIR). We can control the STI by convolving speech with the simulated RIR. Word intelligibility and listening difficulty tests were conducted to compare the proposed method with two conventional methods. The results showed that speech privacy can be protected by controlling STI derived by manipulating the simulated RIR. The results also showed that the proposed method can protect the privacy of conversations as effectively as those other methods can by using lower noise levels and shorter reverberation.

研究分野：音声信号処理

キーワード：音声プライバシー保護 音声伝送指標 変調伝達関数 残響 聴き取りにくさ

1. 研究開始当初の背景

次のような経験をしたことはないだろうか。「薬局のカウンターで薬の服用方法の説明を受けていたときに周辺のお客さんにまで聞こえてしまったこと」や「休憩ブースでの携帯電話の会話が知らずに聞こえてきてしまったこと」などである(図1)。我々は音声プライバシー保護の観点から、音声会話の利用環境や様態について、いつでも細心の注意を払わなければならない。

音声プライバシー保護の代表的な方法は、音環境を半開放の仕切りや防音壁で間仕切りし音の漏えいを防ぐことや、雑音呈示により会話音声をマスクすることで聴き取り難くすることである。しかし、前者は開空間を制約するため利便性を欠くことになるし、後者は音環境の特性を劣化させ騒音暴露を招くことになる。最近では、音声混合音を背景音として情報マスクングを誘発することでプライバシー保護を実現する方法(YAMAHA)や残響音を利用したプライバシー保護(佐藤ら, 神戸大)がある。いずれも情報マスクングを利用するものであり、会話音声の内容を不明瞭にするという意味で絶大な効果がある。しかし、音環境の変化に関係なく特定の処理を行うため、会話音声や音環境の状況変化には対応できない。

一方、音情景解析の概念に基づいた音声プライバシー保護法(赤木&入江, グローリー工業)も提案されている。これは音声と防聴音を融合させることで異聴を招き、会話音声を守るものである。ヒトの聴覚特性に基づいたユニークな研究であるが、会話音声と防聴音との融合による異聴が不自然で違和感があるため、会話者以外(周りに居る人たち)に不快感を持たせるという欠点がある。

研究代表者は、これまでに、室の音響特性を測定せずに観測した音信号から聴き取り難さに関係する音声伝送指標(STI, Speech Transmission Index)の推定法(科研費・挑戦的萌芽, 課題番号 23650086)を確立した。この方法を逆手にとると、室の聴き取り難さに合わせて音声を聴き取り難くすることも原理上は可能である。そのため、この方法を活用することで適切な音声プライバシー保護が可能であると感ぜられる。



図1 音声プライバシー保護

2. 研究の目的

本研究では、室内音響における聴き取り難さと非常に高い相関をもつSTIを操作することで、話者の音声プライバシーを保護する方略を提案する。

近年、情報通信技術の進歩に伴い、いつでもどこでも気軽に音声コミュニケーションをとれるようになった。その一方で、個人のプライバシーに係る会話内容も盗み聞きされる危険性が増してきた。そのため音声プライバシーの保護や音声情報の漏えい防止に関して細心の注意を払わなければならない。従来、マスクングに基づき妨害雑音・音声混合雑音を付与することが主流であるが、単に騒がしいだけで本質的な改善につながっていない。本研究では、STIを陽に操作することで発話内容を不明瞭にしつつ、その加工音声を音環境に溶け込ませることで、音声プライバシー保護を実現する。

表1 音声品質とSTIの関係

Quality	Bad	Poor	Fair	Good	Excel.
STI	0.00	0.30	0.45	0.60	0.71
	~	~	~	~	~
	0.29	0.44	0.59	0.74	0.11

3. 研究の方法

(1) 本研究のアイデア

室内音響特性は、室内を伝送経路(室内音響)と置き換えて、電話による通信と同様に通信経路の物理的な質が情報伝送性能を左右するという枠組みで検討されている。そのため、伝送経路の特性を知るためにインパルス応答を測定し、エネルギー的指標を求めて室内の質の特徴付けを行っている。

代表的なものとして、定常状態からエネルギーが60dB減衰する時間(残響時間 T_{60})や直接音到来までの相対的なエネルギー比(ドイトリヒカイト, D値あるいは D_{50})が利用されている。一方で、入出力間の変調度の比として定義される変調伝達関数(MTF, Modulation Transfer Function)も利用されている。これは音環境での聴き取り難さに対応するSTIの導出で主に利用されている(表1のようにSTIは、室内音響における音声伝送品質を評価するために利用されている)。

我々が音環境の質を主観的に評価するとき、室内インパルス応答(RIR, Room Impulse Response)やエネルギー的指標を予測して音を聴くのだろうか? 答えは否である。心理物理的アプローチを取れば、知覚に関する物理量の変化を検知してそれに伴う質の変化を知覚するものと考えればよく、RIRの推定は不要である。また、聴覚的な意味づけを考えると、これは周波数分析器となる聴覚フィルタバンクの出力における時間包絡線情報に対応しており、時間積分的な関数として音色知覚でも重要な特徴であると考えられる。

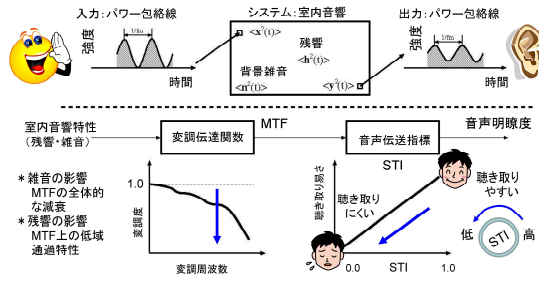


図2 変調伝達関数と音声伝送指標の関係

図2はMTFとSTIの関係を示すものである。MTFは残響等によるパワー包絡線の変調度の減衰を表すものであり、STIはこれに関係して得られる。そのため、音声プライバシー保護を想定する室内でMTFを推定・制御できればSTIも同時に推定・制御可能である。

本研究での挑戦は、このアプローチからSTIを音量ボリュームのように調整することで、聴き取り側の音声明瞭度・了解度を低下させ、結果的に話者の発話内容を不明瞭にすることで音声プライバシー保護を実現することである。原理上、話者の音声に室内インパルス応答とは別のインパルス応答を畳み込むことでSTIを陽に制御可能であるが、(1) どのようなインパルス応答を設計するとよいか、(2) 制御音声をどのように音環境に呈示すればよいか、が未だ不明である。

(2) 提案法の原理

図3に音声プライバシー保護法(提案法)の原理を示す。まず、保護すべきプライベートな会話 $x(t)$ が漏れ、意図しない人が提案法の下で漏れた音を聴くものとする。この漏れた音を $y(t)$ とする。ここで実際の室内のRIRを $h_0(t)$ とすると、実際の漏えい音声は $x(t) * h_0(t)$ となる。次に、プライバシー保護のために処理された音声 $x(t) * h_L(t)$ をだけ遅延させて会話者以外の外部に呈示させる。その結果、意図しない人が聴く音 $y(t)$ は

$$y(t) = h(t) * x(t) \quad (1)$$

と表すことができる。ここで、 $h(t)$ は模擬的なRIRであり、次式のように表される。

$$h(t) = h_0(t) + h_L(t - \tau) \quad (2)$$

室内における元々のSTIを STI_0 とすると、これは $h_0(t)$ から算出される。これに対し、意図しない人が聴く音の模擬的な室内におけるSTIは、 $h(t)$ から同様に算出されるが、 $h_L(t)$ を制御することで如何ようにでもSTIを制御することができる。そこで、意図しない聴取者がいる音環境でのSTIを STI_{tgt} 、 $h_L(t)$ によって求まるSTIを STI_{ext} とすると、実際の STI_0 から提案法によって STI_{ext} を制御することで意図しない聴取者が会話内容を理解できないように STI_{tgt} を制御することができる。以上が提案法の原理である。

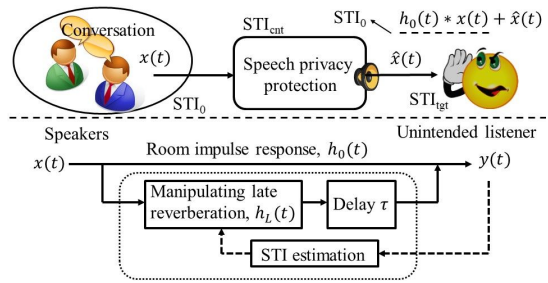


図3 提案法のブロックダイアグラム

(3) RIR/STIの制御法

提案法では、上述の方法に従い、拡張型RIRモデル $h_{ext}(t)$ を利用してSTIを制御する。

まずRIR $h_L(t)$ は次式のように表される。

$$h_L(t) = h_{ext}(t - t_0) \quad (3)$$

ただし

$$h_{ext}(t) = \begin{cases} a \exp(6.9t/T_h) c_h(t) & t < 0 \\ a \exp(-6.9t/T_t) c_h(t) & t \geq 0 \end{cases} \quad (4)$$

である。ここで、 a はゲイン項、 T_h と T_t はRIRの外形を制御するパラメータである。

次に、式(1)の $h(t)$ から得られるMTFを求める。MTF $m(f_m)$ は次式のように表される。

$$m(f_m) = g_0 m_0(f_m) + g_L m_L(f_m, T_h, T_t) \quad (5)$$

ここで g_0 と g_L はそれぞれMTFの係数であり、

$$g_0 = \int h_0^2(t) dt / \int h^2(t) dt \quad (6)$$

$$g_L = \int h_L^2(t) dt / \int h^2(t) dt \quad (7)$$

から得られ、 $h_L(t)$ に対応するMTFは

$$m_L(f_m, T_h, T_t) = \frac{1}{\sqrt{\left(1 + \left(2\pi f_m \frac{T_h}{13.8}\right)^2\right) \left(1 + \left(2\pi f_m \frac{T_t}{13.8}\right)^2\right)}} \quad (8)$$

から得られる。

最後に、上記の関係からSTIを求める。STIは、IEC60268-16 (Sound system equipment - Part 16) によって標準化されており、主に次の手順で求められる。(1) RIRに対する7オクターブ帯域ごとのMTFを求める。(2) 7つのMTFからSNRを算出する。(3) 7つのSNRから伝送指標TIを算出する。(4) TIの平均から変調伝送指標MTIを算出する。(5) MTIの荷重和からSTIを算出する。

図4に、 $\tau = 50$ ms、 $a = \exp(-6.9\tau/T_t)$ としたときの拡張型RIRモデルの外形を、図5に、二つのパラメータ (T_h と T_t) を変数としたときのSTIを示す。このモデルでは、STIを0.23~1.0まで自在に操作することができ、特に $T_h=0.1$ 、 $T_t=10.5$ のときにSTIを0.23に設定することができる。ことがわかる。

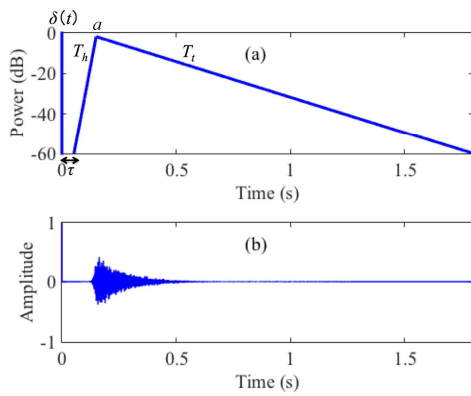


図4 拡張型 RIR モデルの外形

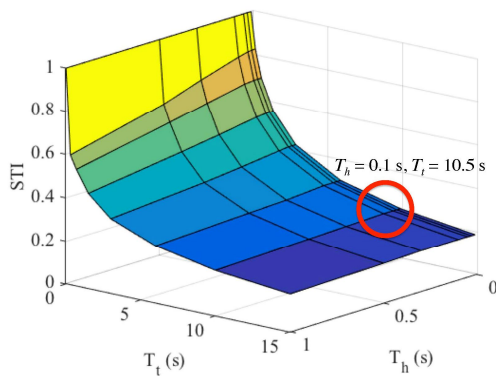


図5 拡張型 RIR モデルの STI

(4) STI のアクティブ制御

前述したように、提案法では STI_{ext} を制御することで所望の STI_{tgt} を得ることができる。ここで制御に必要なパラメータは、遅延時間と拡張型 RIR モデルのパラメータ T_h と T_t である。提案法では、事前に室の STI_0 とプライバシー保護中に意図しない聴取者の場での STI_{tgt} を知らなければならない。これらは、研究代表者が提案した STI ブラインド推定法 (Unoki et al., 2017; 研究業績 1) を利用して、意図しない聴取者の場で観測される $y(t)$ から STI を求めることで実行可能である。特に、3 つの制御パラメータを利用して能動的に STI_{ext} を制御することで、 STI_{tgt} が所望の値になるようにモニタリングしながら制御することができる。

4. 研究成果

(1) 評価方法

式(2)の RIR の第 2 項 $h_L(t)$ を操作して STI を制御することにより、聴取者による音声の聴き取りを制御できるか検討するために、単語了解度試験と聴き取りにくさの聴取実験を行った。

試験用の音声刺激として、親密度別単語了解度試験用データベース (FW07) の男性発話者の親密度 1.0~7.0 の単語を用いた。STI が 0.875, 0.675, 0.525, 0.375, 0.230 になるように RIR を作成し、音声刺激に重畳し残

響音声を作成した。刺激の総数は、STI 5 条件 × 親密度 4 条件 × 20 単語 × 実験数 2 の 800 単語であった。

評価では、 STI_{tgt} が 0.23 (表 1 では “Bad Quality”) であるものとし、元々の室の STI_0 は 1.0 であるものとした。

実験には、正常聴力を有し、日本語を母語とする 23~25 歳の男性 10 名と女性 1 名が参加した。800 単語中、刺激の重複呈示を許さなかった。単語了解度試験では、聴き取った単語をカタカナでキーボード入力させ、その単語正答率を集計し、評価に利用した。

聴き取りにくさの聴取実験では、実験参加者に、聴取音声の印象を、「聴き取りにくくはない」、「やや聴き取りにくい」、「かなり聴き取りにくい」、「非常に聴き取りにくい」の 4 つの評価から 1 つを強制選択させた。ここでは「聴き取りにくくはない」以外の選択数を刺激の総数で除算した値を集計し、その値を評価に利用した。

(2) 結果

提案法が音声プライバシー保護としてどれだけ機能するのか評価するために単語了解度試験と聴き取りにくさの評価を行った。

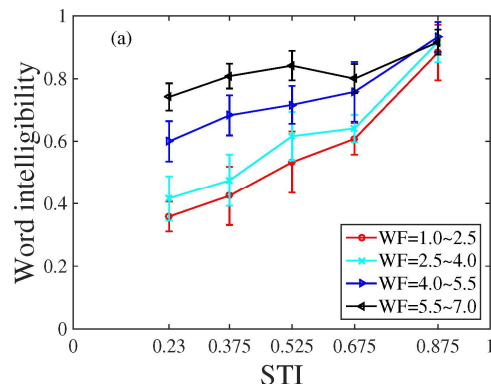


図6 評価結果：単語了解度試験

図 6 に各 STI 条件における親密度別単語了解度試験の結果を示す。図中の WF は親密度を示し、縦棒は標準偏差を示す。図から、STI の低下にともない単語了解度もおおむね低下することがわかった。また、STI が単語了解度に及ぼす影響は、親密度によって異なる傾向があることもわかった。特に、親密度が低いときは、STI の低下にともない単語了解度が低下することがわかった。このことから、単語親密度に注意すれば、STI を操作することで単語了解度を制御できる、つまり音声プライバシーを保護できるといえる。

図 7 に、各 STI 条件における聴き取りにくさの聴取実験の結果を示す。図の表示条件は図 6 と同じである。図から、STI を低下させる、聴き取りにくさが著しく増加し、1.0 に飽和することがわかった。また、この傾向は親密度に依存しないこともわかった。このことから、STI を操作することで聴き取りにくさを制御できるといえる。

以上のことから、STI を操作することで結果的には単語理解度や聞き取りにくさを狙い通りに制御できることがわかった。

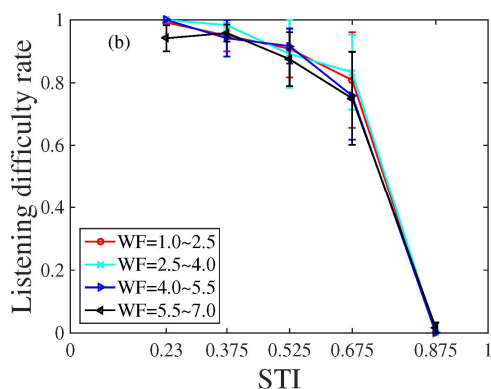


図7 評価結果：聞き取りにくさ

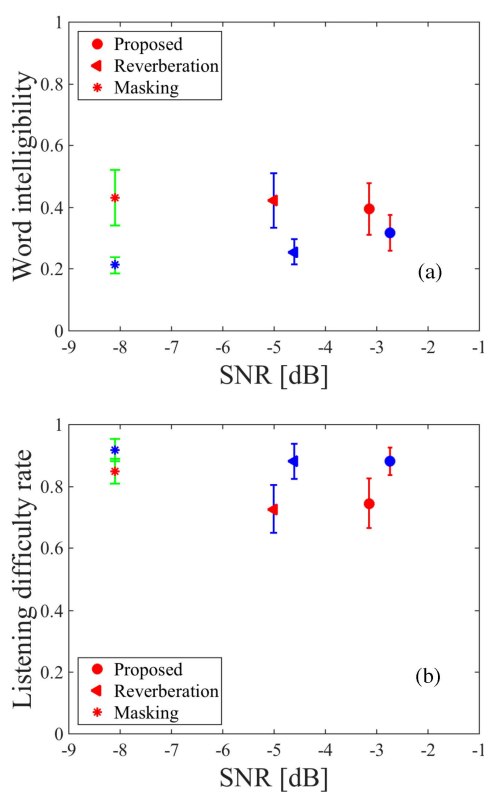


図8 基礎テストにおける比較結果：(a) 単語理解度試験，(b) 聞き取りにくさ

(3) 考察

先の結果から、提案法が音声プライバシー保護として狙いどおり機能することが確認された。しかし従来法と比較してどの程度の効果があるのか不明である。ここでは、音声プライバシー保護の代表的な方法として、残響付与法 (Schroeder の RIR モデルを利用するもの) と、マスキング法 (ピンク雑音を付与するもの) を採用し、同様の条件で評価することにした。残響付与法では、STI が 0.23 になるように残響時間が調整された。マスキング法では STI が 0.23 になるようにピンク雑音の SN 比が調整された。ここで、MTF と

SN 比の換算には次式が利用された。

$$\text{SNR} = 10 \log_{10} m(f_m) / (1 - m(f_m)) \quad (9)$$

提案法に関しても同様に、式(9)を利用して STI から SNR 換算をした。

3つの方法(提案法、残響付与法、マスキング法)を利用したときの評価結果を図8に示す。図のフォーマットは図7と同様であるが横軸は STI から SNR に変更した。また親密度が最も高いものと低いものの2種類のみ掲載した。図から、単語理解度試験と聞き取りにくさの結果が、3つの手法でほとんど変わらないことがわかる。また、親密度の違いによる結果の違いも大きくないことがわかる。

しかし、同程度の結果を得るのに必要な SNR を調べると提案法が最も高く、マスキング法が最も低いこともわかる。このことから、提案法は他の2つの方法よりも音声を妨害するために付与する音のパワーが少なく済んでいる (SNR で 5 dB 程度の改善がみられる) ことがわかる。以上のことから、本研究で提案された方法は従来法よりも優れた音声プライバシー保護法であるといえる。

(4) まとめ

漏えい音声を聴取する人のおかれた音環境の RIR を操作することで、能動的に STI を制御し、結果として漏えい音生の聞き取りにくさの制御も可能であるかどうか検討した。その結果、STI を低くすることで単語理解度を減少させ、さらには聞き取りにくさを増加させることができた。また、STI が単語理解度に及ぼす影響は親密度に依存するが、聞き取りにくさに及ぼす影響は親密度に依存しなかった。以上から、STI を能動的に制御することで、音声のプライバシー保護を実現できることがわかった。

5. 主な発表論文等

〔雑誌論文〕(計6件)

1. Masashi Unoki, Yuta Kashihara, Maori Kobayashi, and Masato Akagi, "Study on method for protecting speech privacy by actively controlling speech transmission index in simulated room," Proc. APSIPA2017, Refereed, Dec. 2017.
DOI: 10.1109/APSIPA.2017.8282212
2. Masashi Unoki, Akikazu Miyazaki, Shota Morita, and Masato Akagi, Method of Blindly Estimating Speech Transmission Index in Noisy Reverberant Environments, Journal of Information Hiding and Multimedia Signal Processing, Refereed, Vol. 8, No. 6, 1430-1445, 2017.
3. Shota Morita, Xugang Lu, Masashi Unoki, and Masato Akagi, Method of Estimating Signal-to-Noise Ratio Based on Optimal

Design for Sub-band Voice Activity Detection, Journal of Information Hiding and Multimedia Signal Processing, Refereed, Vol. 8, No. 6, pp. 1446-1459, 2017.

4. Yuta Kashihara and Masashi Unoki, "Study on Method for Controlling Listening Difficulty by Operating Speech Transmission Index of the Room Acoustics," Proc. 2017 RISP International workshop on Nonlinear Circuits, Communications and Signal Processing (NCSP17), Refereed, pp. 489-492, Guam, USA, March 2017.
5. Daisuke Ishikawa and Masashi Unoki, "Study on method for blindly estimating parameters of the extended stochastic model of room impulse response," Proc. 2017 RISP International workshop on Nonlinear Circuits, Communications and Signal Processing (NCSP17), Refereed, pp. 493-496, Guam, USA, March 2017.
6. Yuta Kashihara and Masashi Unoki, "Study on IIR Implementation for Modulation Transfer Function of Room Impulse Response," J. Signal Processing, Refereed, Vol. 20, No. 4, pp. 201-204, July 2016.
<https://doi.org/10.2299/jsp.20.201>

〔学会発表〕(計10件)

1. 磯山拓都, 鶴木祐史, "変調スペクトルに着目した騒音抑圧法の検討," 電子情報通信学会応用音響研究会, 2018/3/20, 南の美ら花 ホテルミヤヒラ (石垣市).
2. 磯山拓都, 鶴木祐史, "変調スペクトルを用いた騒音低減手法の検討," 日本音響学会 2018 年度春季研究発表会, 2018/3/15, 日本工業大学 (南埼玉郡宮代町).
3. Masashi Unoki, Yuta Kashihara, Maori Kobayashi, Masato Akagi, "Study on speech privacy protection by controlling speech transmission index in simulated room," 日本音響学会聴覚研究会資料, vol. 47, no. 8, H-2017-125, pp. 95-100, 2017/12, オークランド大学 (オークランド).
4. 磯山拓都, 鶴木祐史, "高次統計量を用いた音声・非音声の変調スペクトルの特徴分析に関する検討," 第32回信号処理(SIP)シンポジウム, 2017/11/8, マリオス盛岡地域交流センター (盛岡市).
5. 磯山拓都, 鶴木祐史, "音声・非音声の変調スペクトルの特徴分析に関する検討," 電子情報通信学会電気音響研究会, 2017/10/21, 牛岳温泉リゾート (富山市).
6. 石川大介, 小林まおり, 赤木正人, 鶴木

祐史, "室内インパルス応答のモデル化とその室内音響特性," 日本音響学会 2017 年度春季研究発表会講演論文, 1-P-19, pp. 539-542, 2017/03, 明治大学生田キャンパス(川崎市).

7. 石川大介, 鶴木祐史, "室内インパルス応答モデルのパラメータ推定の検討," 日本音響学会 2017 年度春季研究発表会講演論文, 1-P-20, pp. 543-544, 2017/03, 明治大学生田キャンパス(川崎市).
8. 柏原佑太, 鶴木祐史, "音声伝送指標に着目した音声の聴き取り制御に関する検討," 日本音響学会 2017 年度春季研究発表会講演論文, 3-Q-13, pp. 1431-1432, 2017/03, 明治大学生田キャンパス(川崎市).
9. 鶴木祐史, 石川大介, 柏原佑太, 小林まおり, 赤木正人, "室内インパルス応答のモデル化とその室内音響特性の検討," IEICE Technical Report, EA2016-61, pp. 79-84, 2016/11, 大分コンパルホール(大分市).
10. 柏原佑太, 鶴木祐史, "変調伝達関数モデルを利用した音声伝送指標の検討," IEICE Technical Report, EA2016-33, pp. 13-18, 2016/10, 国民宿舎能登小牧台(七尾市).

〔図書〕(計0件)

〔産業財産権〕

出願状況(計0件)

取得状況(計0件)

〔その他〕

ホームページ等
なし

6. 研究組織

(1)研究代表者

鶴木 祐史 (UNOKI MASASHI)
北陸先端科学技術大学院大学・先端科学技術研究科・教授
研究者番号: 00343187

(2)研究分担者

なし

(3)連携研究者

なし

(4)研究協力者

なし